AD-A034 474 WISCONSIN UNIV MADISON MATHEMATICS RESEARCH CENTER F/G 12/2
GENERALIZED GRADIENTS OF LIPSCHITZ FUNCTIONALS.(U)
OCT 76 F H CLARKE DAAG29-75-C-0024
UNCLASSIFIED MRC-TSR-1687 NL

1 OF 1
ADA034474

END
DATE
FILMED
2 - 77

MRC Technical Summary Report #1687

GENERALIZED GRADIENTS OF LIPSCHITZ FUNCTIONALS

Frank H. Clarke

Mathematics Research Center
University of Wisconsin–Madison
610 Walnut Street
Madison, Wisconsin 53706

October 1976

(Received August 10, 1976)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

GENERALIZED GRADIENTS OF LIPSCHITZ FUNCTIONALS

Frank H. Clarke

Technical Summary Report # 1687
October 1976

ABSTRACT

The purpose of this article is threefold: (i) to present in a unified fashion the theory of generalized gradients, whose elements are at present scattered in various sources; (ii) to give an account of the ways in which the theory has been applied; (iii) to prove new results concerning generalized gradients of summation functionals, pointwise maxima, and integral functionals on subspaces of L^∞. These last-mentioned formulas are obtained with an eye to future applications in the calculus of variations and optimal control (one such is given in the final section). Their proofs can be regarded as applications of the existing theory of subgradients of convex functionals as developed by Rockafellar, Ioffe and Levin, Valadier, and others.

AMS(MOS) Subject Classification - 46G05, 26A24, 26A27

Key Words - Locally Lipschitz functions, subdifferential, generalized derivatives, generalized gradients

Work Unit #1 - Applied Analysis

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

GENERALIZED GRADIENTS OF LIPSCHITZ FUNCTIONALS

Frank H. Clarke*

The purpose of this article is threefold: (i) to present in a unified fashion the theory of generalized gradients, whose elements are at present scattered in various sources (§1); (ii) to give an account of the ways in which the theory has been applied (§2); (iii) to prove new results concerning generalized gradients of summation functionals (§3), pointwise maxima (§4) and integral functionals on subspaces of L^{∞} (§5). These last-mentioned formulas are obtained with an eye to future applications in the calculus of variations and optimal control (one such is given in §6). Their proofs can be regarded as applications of the existing theory of subgradients of convex functionals as developed by Rockafellar, Ioffe and Levin, Valadier and others (see [19] [26] for references).

1. Definition. Basic properties

Let U be an open subset of a Banach space X, and let a function $f: U \to R$ be given. We shall suppose that f is Lipschitz on U; i.e. that for some constant K, for all u_1 and u_2 in U, we have

$$|f(u_1) - f(u_2)| \le K\|u_1 - u_2\| . \tag{1.1}$$

---

* Department of Mathematics, The University of British Columbia, Vancouver, B.C., Canada V6T 1W5

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

Let us now fix a point x in U and any point v in X.

Definition 1. The generalized directional derivative of f at x in the direction v, denoted $f^{\circ}(x;v)$, is given by

$$f^{\circ}(x;v) = \limsup_{\substack{y \to x \\ \lambda \downarrow 0}} [f(y+\lambda v) - f(y)]/\lambda \quad . \tag{1.2}$$

(Here of course y belongs to X and λ to $(0,\infty)$.)

Note that this definition does not presuppose the existence of any limit, and that in view of (1.1), $f^{\circ}(x;v)$ is a finite number for all v in X. The following observation is most important:

Lemma. The function $v \to f^{\circ}(x;v)$ is positively homogeneous and subadditive, and satisfies

$$f^{\circ}(x;v) \leq K\|v\| \quad .$$

Proof: The homogeneity and the inequality are immediate consequences of the definition. As for the subadditivity, let v and w in X be given. Then

$$\begin{aligned} f^{\circ}(x;v+w) &= \limsup_{\substack{y \to x \\ \lambda \downarrow 0}} [f(y + \lambda v + \lambda w) - f(y)]/\lambda \\ &\leq \limsup_{\substack{y \to x \\ \lambda \downarrow 0}} [f(y + \lambda v + \lambda w) - f(y + \lambda w)]/\lambda \\ &\quad + \limsup_{\substack{y \to x \\ \lambda \downarrow 0}} [f(y + \lambda w) - f(y)]/\lambda \\ &= f^{\circ}(x;v) + f^{\circ}(x;w) \, . \end{aligned}$$

Q.E.D.

In view of the above lemma, it follows from the Hahn-Banach theorem [12, p. 62] that there exists at least one linear functional $\zeta: X \to R$ satisfying

(1.3) $$f^{\circ}(x;v) \geq \langle v, \zeta \rangle \quad \text{for all } v \text{ in } X\,;$$

it is a further consequence of the lemma that ζ is *continuous*. Thus ζ belongs to X^*.

(As usual, X^* denotes the (continuous) dual of X and $\langle \cdot\,, \cdot \rangle$ is the duality pairing between X and X^*.)

<u>Definition 2.</u> The generalized gradient of f at x, denoted $\partial f(x)$, is the (nonempty) set of all ζ in X^* satisfying (1.3).

We now proceed to discuss some of the fundamental results in the calculus of generalized gradients.

1. Nature of ∂f [9, Proposition 1].

$\partial f(x)$ is a nonempty convex subset of X^*. It is closed in the strong topology of X^* and bounded by K; thus $\partial f(x)$ is w^*-compact.

2. $f^{\circ}(x;\cdot)$ is the support function of $\partial f(x)$ [9, Proposition 1].

This means that for any v in X, we have

$$f^{\circ}(x;v) = \max\{\langle v, \zeta \rangle : \zeta \in \partial f(x)\}\ .$$

3. $\partial f(\cdot)$ is w^*-upper semicontinuous [9, Proposition 7].

I.e., if $\zeta_i \in \partial f(x_i)$ where $x_i \to x$ in X and $\zeta_i \to \zeta$ in $X^*(w^*)$, then $\zeta \in \partial f(x)$.

4. $f^{\circ}(\cdot\,;\cdot)$ is upper semicontinuous.

Let $x_i \to x$ and $v_i \to v$ in X. For each i there exist h_i in X and λ_i in $(0,1)$ such that $\|h_i\| + \lambda_i$ is less than $1/i$ and

$$f^{\circ}(x_i;v_i) \leq [f(x_i + h_i + \lambda_i v_i) - f(x_i + h_i)]/\lambda_i + 1/i .$$

But the limit superior of the right side of this inequality (as $i \to \infty$) is easily seen to be no greater than $f^{\circ}(x;v)$. We obtain

$$\limsup_{i \to \infty} f^{\circ}(x_i;v_i) \leq f^{\circ}(x;v) . \qquad \text{Q.E.D.}$$

5. $\partial(-f)(x) = -\partial f(x)$.

This follows from the observation that $(-f)^{\circ}(x;v) = f^{\circ}(x;-v)$.

6. If f attains a local minimum or maximum at x, then $0 \in \partial f(x)$ [9, Proposition 6].

7. The mean value property (G. Lebourg [20]).

If x and y are distinct points of X then there is a point z in the open line segment between x and y such that

$$f(y) - f(x) \in \langle y-x, \partial f(z) \rangle .$$

(We assume here that all points in question lie in the set U upon which f is Lipschitz.)

8. $\partial f(x)$ when $X = R^n$ [4].

In this case, $\partial f(x)$ is the set $\mathrm{co}\{\lim_{i \to \infty} \nabla f(x_i) : x_i \to x\}$. That is (f being differentiable a.e. by Rademacher's theorem), we consider all sequences x_i converging to x such that f is differentiable at x_i and the indicated limit exists. The convex hull of these limits is $\partial f(x)$.

This is equivalent to saying that ∂f is the minimal upper-semicontinuous convex-valued multifunction containing the derivative when it exists.

9. If f is convex, $\partial f(x)$ coincides with the subdifferential in the sense of convex analysis [9, Proposition 3]. In view of article 5, a similar statement holds if f is concave.

10. If f admits a continuous Gâteaux derivative Df, then $\partial f(x) = \{Df(x)\}$ [9, Proposition 4].

When $X = R^n$, $\partial f(x)$ reduces to a singleton set $\{\zeta\}$ iff f is strongly differentiable at x and $Df(x) = \zeta$ (f is said to be strongly differentiable at x if

$$\lim_{\substack{y \to x \\ v \to 0}} [f(y+v) - f(y) - Df(x)(v)]/\|v\| = 0 ;$$

Bourbaki [2] uses the term "strictement dérivable".) Lebourg [21, Theorem 2.1] gives a similar result in the infinite dimensional case. This shows that a differentiable function may have a generalized gradient containing points other than the derivative. The latter is always contained in ∂f whenever it exists.

11. Generalized gradients of sums.

The inclusion $\partial(f+g) \subset \partial f + \partial g$ was established in [4]. A more general study (including continuous sums and conditions for equality) appears in §3.

12. Pointwise maxima.

One can relate the generalized gradient of $x \to \max_t f(t, x)$ to the generalized gradients of the functions $x \to f(t, x)$; see §4.

Definition 3. We shall say that f is regular at x if for every v in X the usual one-sided directional derivative

$$f'(x;v) = \lim_{\lambda \downarrow 0} [f(x+\lambda v) - f(x)]/\lambda$$

exists, and satisfies $f'(x;v) = f^{\circ}(x;v)$. Convex functions and continuously differentiable functions are regular at every point (and certain quasidifferentiable functions; see article 16).

13. Chain rule I.

Let g map X to another Banach space Y, and suppose that g is continuously Gâteaux differentiable. Let $h : Y \to R$ be Lipschitz. Then, if $f: X \to R$ is given by $f = h \circ g$, we have

$$\partial f(x) \subset \partial h(g(x)) \circ Dg(x) . \tag{1.4}$$

Equality holds if either h (or $-h$) is regular at $g(x)$ or $Dg(x)$ is surjective.

Proof: Since the right side of (1.4) is convex and w^*-compact, it suffices to prove that, for any v in X,

$$f^{\circ}(x;v) \leq \max\{\langle Dg(x)\, v, \zeta\rangle : \zeta \in \partial h(g(x))\}$$

(see [9, Proposition 2]). Now any expression of the form

$$[f(y+\lambda v) - f(y)]/\lambda$$

is equal to $\langle g(y+\lambda v) - g(y), \zeta\rangle/\lambda$ for some $\zeta \in \partial h(z)$, where z is in the interval between $g(y)$ and $g(y+\lambda v)$ (this is an application of article 7). Further, by the mean value theorem,

$$[g(y+\lambda v) - g(y)]/\lambda = Dg(w)$$

for some w in the interval between y and $y+\lambda v$. As y converges to x and λ to 0 in the above, we can (by taking an appropriate subsequence) assume that ζ converges (w^*) to an element ζ_0 of $\partial h(g(x))$ (in light of articles 1 and 3). Of course $Dg(w)$ converges to $Dg(x)$. We obtain

$$\limsup_{\substack{y \to x \\ \lambda \downarrow 0}} [f(y+\lambda v) - f(y)]/\lambda \leq \langle Dg(x)v, \zeta_0 \rangle ,$$

which yields the desired inequality.

Now suppose that h is regular at $g(x)$. Then by article 2

$$\begin{aligned}
&\max \{\langle Dg(x)v, \zeta\rangle : \zeta \in \partial h(g(x))\} \\
&\quad = h^{\circ}(g(x); Dg(x)v) = h'(g(x); Dg(x)v) \\
&\quad = \lim_{\lambda \downarrow 0} [h(g(x) + \lambda Dg(x)v) - h(g(x))]/\lambda \\
&\quad = \lim_{\lambda \downarrow 0} [h \circ g(x+\lambda v) - h \circ g(x)]/\lambda \\
&\quad = f'(x;v) \leq f^{\circ}(x;v) = \max\{\langle v, \zeta\rangle : \zeta \in \partial f(x)\} .
\end{aligned}$$

This implies

$$\partial h(g(x)) \circ Dg(x) \subset \partial f(x) , \tag{1.5}$$

as required.

Finally, suppose that $Dg(x)$ is surjective. Then by the interior mapping theorem, g maps every neighborhood of x to a neighborhood of $g(x)$. We deduce

$$
\begin{aligned}
&\max\{\langle Dg(x)v, \zeta\rangle : \zeta \in \partial h(g(x))\} \\
&\qquad = h^{\circ}(g(x); Dg(x)v) \\
&\qquad = \limsup_{\substack{z \to g(x) \\ \lambda \downarrow 0}} [h(z + \lambda Dg(x)v) - h(z)]/\lambda \\
&\qquad = \limsup_{\substack{y \to x \\ \lambda \downarrow 0}} [h(g(y) + \lambda Dg(x)v) - h(g(y))]/\lambda \\
&\qquad = \limsup_{\substack{y \to x \\ \lambda \downarrow 0}} [h(g(y + \lambda v)) - h(g(y))]/\lambda \\
&\qquad = f^{\circ}(x;v) \ .
\end{aligned}
$$

We deduce (1.5) as before. Q.E.D.

14. Chain rule II.

Let $g: X \to R$ and $h: R \to R$ be Lipschitz. Then if $f: X \to R$ is defined by $f = h \circ g$,

$$\partial f(x) \subset \text{co}[\partial h(g(x))\ \partial g(x)] \ .$$

Furthermore, if h is C^1, or if h (or $-h$) is regular at $g(x)$ and g is continuously Gâteaux differentiable, the symbol "co" is superfluous and equality holds.

The proof is much like that of chain rule I, and is therefore omitted.

15. Partial generalized gradients.

Let $f: X \times Y \to R$ be Lipschitz. Then it is natural to denote, for each x, the generalized gradient of the function $y \to f(x, y)$ by $\partial_y f(x, y)$; similarly we define $\partial_x f(x, y)$. There exists a function f (with $X = Y = R$) such that neither of the sets $\partial f(x, y)$, $\partial_x f(x, y) \times \partial_y f(x, y)$ is contained in the other. We have, however, the following: if f (or $-f$) is regular at (x, y) then

$$\partial f(x, y) \subset \partial_x f(x, y) \times \partial_y f(x, y) \quad .$$

<u>Proof:</u> Let (ζ, φ) belong to $\partial f(x, y)$. It suffices to prove that ζ belongs to $\partial_x f(x, y)$, which in turn is equivalent to the condition that, for all v,

$$f'_x(x, y; v) = f^\circ_x(x, y; v) \geq \langle v, \zeta \rangle \quad .$$

But we have

$$f'_x(x, y; v) = f'(x, y; v, 0) = f^\circ(x, y; v, 0) \geq \langle (v, 0), (\zeta, \varphi) \rangle$$

$$= \langle v, \zeta \rangle \quad . \qquad \text{Q.E.D.}$$

A further result in this direction is the following: if X and Y are finite-dimensional and f is convex as a function of x alone, then $(\zeta, \varphi) \in \partial f(x, y)$ implies $\zeta \in \partial_x f(x, y)$. This may be proven by means of article 8.

16. Quasidifferentiable functions (Pshenichnyi [25]).

A function f admitting one-sided directional derivatives $f'(x;v)$ in the usual sense is said to be quasi-differentiable if there is a convex w^*-compact subset $M(x)$ of X^* such that for each x, for every v in X, (1.6) $$f'(x;v) = \max\{\langle v, \zeta\rangle : \zeta \in M(x)\} .$$

We shall prove the following: if f is quasi-differentiable and the multifunction M is upper semicontinuous, then f is locally Lipschitz and regular, and $M(x) = \partial f(x)$ (the upper semicontinuity is with respect to the strong topology on X^*).

Proof: Fix x and $\varepsilon > 0$. For all y near x, t near 0 and v in B, we have $M(y+tv) \subset M(x) + \varepsilon B^*$ (where B, B^* are the unit balls in X, X^*). This implies that for all such y, t, v we have $f'(y+tv;v) \leq f'(x;v) + \varepsilon$. If we fix any such y, v and define $g(t) = f(x+tv)$ for $0 \leq t \leq T$ (T independent of y and v), it follows that g has a uniformly bounded upper right Dini derivate, which is known to imply that g is Lipschitz [17, (17.23)]. Thus we may write, for any $0 < \tau < T$,

$$[f(y+\tau v) - f(y)]/\tau = \frac{1}{\tau}\int_0^\tau \frac{d}{dt}\, g(t)dt$$

$$= \frac{1}{\tau}\int_0^\tau f'(y+tv;v)dt$$

$$\leq f'(x;v) + \varepsilon .$$

It follows that f is Lipschitz in a neighborhood of x, and we deduce as well $f^{\circ}(x;v) \leq f'(x;v) + \varepsilon$. Since ε is arbitrary, we conclude that f° and f' coincide (i.e. f is regular). That $\partial f(x)$ and $M(x)$ are equal now follows immediately from (1.6) and article 2. Q.E.D.

17. Evaluation functions.

Let Φ be a Banach space of functions from a space T to a

Banach space X such that, that for a certain t in T, the mapping $\varphi \to \varphi(t)$ from Φ to R is continuous. If $f: X \to R$ is a given Lipschitz function, define $F: \Phi \to R$ by

$$F(\varphi) = f(\varphi(t)) \quad .$$

Then F is Lipschitz, and for every element ζ of $\partial F(\bar{\varphi})$ there exists an element ζ_t of $\partial f(\bar{\varphi}(t))$ such that

$$\langle v, \zeta \rangle = \langle v(t), \zeta_t \rangle \quad \text{for all } v \text{ in } \Phi \quad .$$

Proof: Since $\pi(\varphi) = \varphi(t)$ is a continuous linear functional, it is Lipschitz, and hence so is F. We have $F = f \circ \pi$, and $D\pi(\bar{\varphi}) = \pi$. The result *now follows from chain rule I.* Q. E. D.

2. *Related work.*

Generalized gradients were introduced in the author's thesis [3] and in [4] for the case $X = R^n$; the infinite-dimensional case was broached in [9]. They have been used in applications to the calculus of variations [5] [7] [11], optimal control [6] [10], flow-invariant sets [1] [4], the inverse function theorem [8] and mathematical programming [9]. Numerical algorithms employing generalized gradients have been developed by A. A. Goldstein [14], A. Feuer [13] and R. Mifflin [22]. G. Lebourg [20] [21] obtained the mean value theorem of §1 and used generalized gradients as a tool to investigate generic differentiability properties of locally Lipschitz functions on Banach spaces. Hiriart-Urruty [18] applies the theory to mathematical programming, and

B. S. Morduhovic [23] to control problems. L. Thibault [28] extends property 8 of §1 to separable Banach spaces by means of the Haar derivative. In [24] B. Pourciau discusses the properties of generalized Jacobians (the extension of generalized gradients to vector-valued functions) introduced in [8] (see also Sweetser [27]).

Among the many notions that generalize the concept of the derivative, there are two (introduced subsequent to generalized gradients) that are closely related to the present work. These are "derivate containers" and "screens", defined by J. Warga [29] and H. Halkin [15] respectively. It can be shown that when $X = R^n$, these concepts are more general than that of generalized gradient. In infinite dimensions however, these concepts are not applicable to all locally Lipschitz functions (in contrast to generalized gradients) because of the fact that, not being intrinisically defined, they require that f be uniformly approximated by continuously differentiable functions. As pointed out to us by J. Warga, this is not always possible (an example is the norm on the space $C[0,1]$).

3. Generalized gradients of summation functionals.

Let $(T, \mathcal{T}, \mu)$ be a positive measure space, and let U be an open subset of a Banach space X. We suppose given a function f: $T \times U \to R$, and we assume that for some $k \in L^1(T, R)$ (the space of integrable functions from T to R) we have, for all t in T and u_1, u_2 in U,

$$(3.1) \qquad |f(t,u_1) - f(t,u_2)| \le k(t)\|u_1 - u_2\| \ .$$

Finally we suppose that for each x in U, $t \to f(t,x)$ is measurable, and we define $F: X \to R$ as follows:

$$F(x) = \int_T f(t,x)\,\mu(dt) \ ,$$

whenever this integral is defined.

Theorem 1. Suppose that at least one of the following conditions is satisfied:

(a) T is countable; or

(b) X is separable; or

(c) T is a separable metric space, μ is a regular measure, and the mapping $t \to \partial_x f(t,x)$ is upper semicontinuous (w^*) for each x.

If $F(x)$ is defined for some point x in U, then F is defined and Lipschitz in a neighborhood of x, and

$$(*) \qquad \partial F(x) \subset \int_T \partial_x f(t,x)\,\mu(dt) \ ,$$

by which we mean that to every ζ in $\partial F(x)$ there corresponds a mapping $t \to \zeta(t)$ from T to X^* such that $t \to \langle v, \zeta(t)\rangle$ belongs to $L^1(T,R)$ for every v in X,

$$\langle v, \zeta\rangle = \int_T \langle v, \zeta(t)\rangle\,\mu(dt)$$

for every v in X, and

$$\zeta(t) \in \partial_x f(t,x) \qquad \mu\text{-a.e.}$$

If $f(t,\cdot)$ is regular (see Definition 3, §1) for each t, then F is regular and equality holds in (*).

Proof: That F is defined and Lipschitz near x is an immediate consequence of the hypotheses, in particular (3.1).

Let ζ belong to $\partial F(x)$. Then, for any v in X,

$$F^{\circ}(x;v) = \limsup_{\substack{y \to x \\ \lambda \downarrow 0}} \int_T [f(t, y+\lambda v) - f(t,y)]/\lambda \quad \mu(dt) .$$

Condition (3.1) allows us to invoke Fatou's Lemma to bring the lim sup under the integral sign and deduce:

$$(3.2) \qquad \int_T f_x^{\circ}(t,x;v)\,\mu(dt) \geq F^{\circ}(x;v) \geq \langle v, \zeta \rangle ,$$

(the last inequality being a consequence of article 2, §1). Let us define $\hat{f}(t,v) = f_x^{\circ}(t,x;v)$. It follows that $\hat{f}$ is convex in v, and that, for all t in T and v in X,

$$|\hat{f}(t,v)| \leq k(t) .$$

We claim that $\hat{f}$ is measurable as a function of t. If (a) holds this is automatic, whereas if (c) holds it follows from the easily proven fact that $f_x^{\circ}(t,x;v)$ is upper semicontinuous with respect to t (use §1, article 2). It remains to consider case (b). Let $\{d_n\}$ be a countable dense subset of X. It follows that $f^{\circ}(t,x;v)$ is equal to

$$\limsup_{\substack{d_n \to 0 \\ \lambda \downarrow 0 \\ \lambda \text{ rational}}} [f(t, x+d_n+\lambda v) - f(t,x+d_n)]/\lambda .$$

But for each n and λ this last expression is measurable by hypothesis; it follows that the "countable lim sup" defines a measurable function of t.

If we now define a convex continuous function $\hat{F}: X \to R$ via

$$\hat{F}(v) = \int_T \hat{f}(t, v)\, \mu(dt) \quad ,$$

then (3.2) asserts that ζ belongs to the subdifferential of $\hat{F}$ at 0. We now apply [19, Theorem 1, p. 8] in cases (a) and (b), and [19, Theorem 1, p. 13] in case (c), to deduce exactly the conclusion of the theorem regarding ζ (notice that $\partial \hat{f}(t, 0) = \partial_v f_x^{\circ}(t, x; 0) = \partial_x f(t, x)$ by Definition 2).

Now let us suppose that $f(t, \cdot)$ is regular. From the dominated convergence theorem it follows that

$$\begin{aligned} F^{\circ}(x;v) \geq F'(x;v) &= \int_T f'_x(t, x;v)\, \mu(dt) \\ &= \int_T f_x^{\circ}(t, x;v)\, \mu(dt) \\ &\geq F^{\circ}(x;v) \quad , \end{aligned}$$

the last inequality having been established earlier. This shows that F is regular. Finally, let $\zeta \,(= \int_T \langle \cdot, \zeta(t)\rangle\, \mu(dt))$ be an element of the right side of (*). Then (since $\zeta(t) \in \partial_x f(t, x)$ μ-a.e.)

$$\begin{aligned} F^{\circ}(x;v) &= \int_T f_x^{\circ}(t, x;v)\, \mu(dt) \\ &\geq \int_T \langle v, \zeta(t)\rangle\, \mu(dt) = \langle v, \zeta\rangle \quad . \end{aligned}$$

This implies that ζ belongs to $\partial F(x)$. Q.E.D.

4. Generalized gradients of pointwise maxima.

Let T be a sequentially compact topological space, and let $f: T \times U \to R$ satisfy

$$|f(t,u_1) - f(t,u_2)| \leq K\|u_1 - u_2\|$$

for all t in T and u_1, u_2 in U, where U is an open subset of X. We suppose that $f(\cdot, x)$ is upper semicontinuous for every x in U, and we set

$$F(x) = \max_{t \in T} f(t,x) \quad .$$

It is easy to prove that F is Lipschitz on U.

<u>Theorem 2.</u> Suppose that at least one of the following holds:

(a) X is separable; or

(b) T is metrizable (in particular, if T is separable),

and suppose that $\partial_x f(t,x)$ is upper semicontinuous (w^*) in (t,x).

Then for any x in U,

$$(*) \qquad \partial F(x) \subset \bigcup \{ \int_T \partial_x f(t,x)\,\mu(dt): \ \mu \in P[T(x)] \} \ ,$$

where $T(x)$ is the (closed) set

$$\{t \in T: \ F(x) = f(t,x)\}$$

and $P[T(x)]$ is the set of probability Radon measures supported on $T(x)$. by this we mean that to every ζ in $\partial F(x)$ there correspond an element μ of $P[T(x)]$ and a mapping $\zeta(t)$ from T to X^* such that for every v in X, $t \to \langle v, \zeta(t) \rangle$ belongs to $L^1(T,R)$ (with respect to μ) and

$$\langle v, \zeta \rangle = \int_T \langle v, \zeta(t) \rangle \, \mu(dt) \ .$$

Furthermore, if $f(t, \cdot)$ is regular for each t (see Definition 3), then F is regular and equality holds in (*) .

Proof: We begin by proving:

$$F^{O}(x;v) \leq \max_{t \in T(x)} f_x^{O}(t, x;v) \ . \tag{4.1}$$

Note first that $f_x^{O}(t, x;v)$ is upper semicontinuous and $T(x)$ is compact, so that the notation "max" is justified. For any y near x and λ near 0, we have

$$[F(y+\lambda v) - F(y)]/\lambda \leq [f(t, y+\lambda v) - f(y)]/\lambda \ ,$$

for any t in $T(y+\lambda v)$. Further (§1, article 7), there is a point $z_{y\lambda}$ in the interval between y and $y+\lambda v$ and a point $\zeta_{y\lambda}$ belonging to $\partial_x f(t, z_{y\lambda})$ such that

$$[f(t, y+\lambda v) - f(y)]/\lambda = \langle v, \zeta_{y\lambda} \rangle \ .$$

If we now assume that we have a sequence of such (y, λ) converging to $(x, 0)$, we can pick a subsequence such that $\zeta_{y\lambda}$ converges w^* to a point ζ_0 and the points t converge to some t_0 in T . It follows that t_0 belongs to $T(x)$, and that ζ_0 belongs to $\partial_x f(t_0, x)$. From this argument we can conclude the validity of 4.1, since $\langle v, \zeta_0 \rangle \leq f_x^{O}(t_0, x;v)$.

Now let ζ belong to $\partial F(x)$. Then (4.1) implies

$$\max_{t \in T(x)} \hat{f}(t,v) \geq \langle v, \zeta \rangle \quad \text{for all } v \ ,$$

where $\hat{f}(t,v) = f_x^o(t,x;v)$. Because $\hat{f}$ is upper semicontinuous in t and finite convex in v, we can apply [19, Theorem 2, p. 33] if (a) holds, or [19, Theorem 3, p. 34] if (b) holds, to conclude that ζ belongs to the subdifferential at 0 of the function $v \to \max_{t \in T(x)} \hat{f}(t,v)$ and hence has exactly the form indicated in the theorem (we use here the fact that $\partial_v \hat{f}(t,0) = \partial_x f(t,x)$).

Now let us suppose that each $f(t,\cdot)$ is regular, and let us set

$$\underline{F}'(x;v) = \liminf_{\lambda \downarrow 0} [F(x+\lambda v) - F(x)]/\lambda \ .$$

We certainly have $\underline{F}'(x;v) \leq F^o(x;v)$. In order to prove that $F'(x;v)$ exists and equals $F^o(x;v)$ (i.e. that F is regular) it suffices to prove the opposite inequality.

To this end, note that for any $\lambda > 0$, for any t in $T(x)$, we have

$$[F(x+\lambda v) - F(x)]/\lambda \geq [f(t,x+\lambda v) - f(t,x)/\lambda \ .$$

Taking the lim inf of both sides, we obtain

$$\underline{F}'(x;v) \geq f_x'(t,x;v) = f_x^o(t,x;v) \ .$$

Since this is true for any t in $T(x)$, we deduce

$$\underline{F}'(x;v) \geq \max_{t \in T(x)} f_x^o(t,x;v) \geq F^o(x;v) \ ,$$

in light of (4.1). This completes the proof that F is regular.

We now show that equality holds in (*). Let ζ belong to the right side of (*), with μ and $\zeta(t)$ characterizing ζ as described there. Then, for any v in X,

$$
\begin{aligned}
\langle v, \zeta \rangle &= \int_T \langle v, \zeta(t) \rangle \ \mu(dt) \\
&= \int_T f_x^{\circ}(t, x; v) \ \mu(dt) = \int_T f'_x(t, x; v) \ \mu(dt) \\
&= \lim_{\lambda \downarrow 0} \{ \int_T f(t, x+\lambda v) \ \mu(dt) - \int_T f(t, x) \ \mu(dt) \} / \lambda \\
&\leq \limsup_{\lambda \downarrow 0} \{ \int_T F(x+\lambda v) \ \mu(dt) - \int_T F(x) \ \mu(dt) \} / \lambda \\
&= \limsup_{\lambda \downarrow 0} [F(x+\lambda v) - F(x)] / \lambda = F^{\circ}(x; v) \ .
\end{aligned}
$$

Thus for any v in X we have

$$F^{\circ}(x;v) \geq \langle v, \zeta \rangle \quad ,$$

i.e. ζ belongs to $\partial F(x)$. Q. E. D.

Remark. Important special cases in which the upper semicontinuity of $\partial_x f(t, x)$ and regularity of $f(t, \cdot)$ are present occur when f is continuous in t and convex in x, or when f admits a (jointly) continuous derivative $D_x f$.

5. Integral functionals on subspaces of $L^{\infty}(T, X)$.

In this section we suppose that $(T, \mathcal{J}, \mu)$ is a σ-finite positive measure space and X a separable Banach space. $L^{\infty}(T, X)$ denotes the space of (measurable) essentially bounded functions $\varphi: T \to X$. We suppose given a function $f: T \times U \to R$ (where U is an open subset of X) with the following property: for some $\varepsilon > 0$, there is a function $k \in L^1(T, R)$ such that for all t in T, for all u_1, u_2 in an ε-neighborhood of U, we have

$$|f(t, u_1) - f(t, u_2)| \leq k(t) \|u_1 - u_2\| .$$

We assume that f is measurable as a function of t.

Finally, we assume that for at least one $\varphi \in L^\infty(T, U)$ the integral

$$F(\varphi) = \int_T f(t, \varphi(t)) \, \mu(dt)$$

is defined (finitely). It follows that the integral is defined for all φ in $L^\infty(T, U)$, and that F is locally Lipschitz as a function from $L^\infty(T, U)$ to R.

Now let a closed subspace S of $L^\infty(T, X)$ be given, and let us consider F as a function from S to R (only the values of F on $S \cap L^\infty(T, U)$ will be relevant).

Theorem 3. If $F: S \to R$ is as defined above, then for any point φ in $L^\infty(T, U) \cap S$ we have

$$(*) \qquad \partial F(\varphi) \subset \int_T \partial_x f(t, \varphi(t)) \, \mu(dt) ,$$

by which we mean the following: to every ζ in $\partial F(\varphi)$ there corresponds a mapping $t \to \zeta(t)$ from T to X^* such that $\zeta(t)$ belongs to $\partial_x f(t, \varphi(t))$

μ-a.e. and such that, for all β in S, $t \to \langle \beta(t), \zeta(t) \rangle$ belongs to $L^1(T, R)$ and

$$\langle \beta, \zeta \rangle = \int_T \langle \beta(t), \zeta(t) \rangle \mu(dt) .$$

Furthermore, if for each t, $f(t, \cdot)$ is regular (see Definition 3), then F is regular and equality holds in (*).

<u>Proof:</u> Let ζ belong to $\partial F(\varphi)$, and let β be any element of S. Then for λ small, $\varphi(t) + \lambda\beta(t)$ belongs to the ε-neighborhood of U for all t, and by Fatou's Lemma

$$\int_T f_x^o(t, \varphi(t); \beta(t)) \mu(dt) \geq F^o(\varphi;\beta) \geq \langle \beta, \zeta \rangle .$$

If we define $\hat{f}(t, x) = f_x^o(t, \varphi(t); x)$, then $\hat{f}$ is continuous and convex in x and measurable in t (the latter fact follows just as it did in case (b) of Theorem 1). Consequently $\hat{f}$ is a "normal convex integrand".

The inequality above says that ζ belongs to the subdifferential at 0 of the function (on S)

$$\beta \to \int_T \hat{f}(t, \beta(t)) \mu(dt) .$$

The requisites of [19, Theorem 2, p. 22] are present, so that we deduce the existence of a mapping $t \to \zeta(t)$ such that for every β in S, $t \to \langle \beta(t), \zeta(t) \rangle$ belongs to $L^1(T, R)$ and

$$\langle \beta, \zeta \rangle = \int_T \langle \beta(t), \zeta(t) \rangle \mu(dt) ,$$

and such that

$$\zeta(t) \in \partial \hat{f}(t, 0) \quad \mu\text{-a.e.}$$

But $\partial \hat{f}(t, 0)$ is equal to $\partial_x f(t, \varphi(t))$ by §1, article 2. This completes the proof of the first part.

Now suppose that $f(t, \cdot)$ is regular. If we set

$$\underline{F}'(\varphi ; \beta) = \liminf_{\lambda \downarrow 0} [F(\varphi + \lambda\beta) - F(\varphi)]/\lambda \ ,$$

we have (invoking Fatou's lemma)

$$F^{O}(\varphi;\beta) \geq \underline{F}'(\varphi;\beta) \geq \int_T f'_x(t, \varphi(t); \beta(t))\, \mu(dt)$$

$$= \int_T f^{O}_x(t, \varphi(t); \beta(t))\, \mu(dt) \geq F^{O}(\varphi;\beta) \ .$$

It follows that F is regular.

Now let ζ belong to the right side of (*), with representation $\zeta(t)$. Then

$$\langle \beta, \zeta \rangle = \int_T \langle \beta(t), \zeta(t) \rangle\, \mu(dt)$$

$$\leq \int_T f^{O}_x(t, \varphi(t); \beta(t))\, \mu(dt)$$

$$= \int_T f'_x(t, \varphi(t); \beta(t))\, \mu(dt) = F^{O}(\varphi;\beta) \ ,$$

the last equality being a consequence of the preceding calculations. This implies that ζ belongs to $\partial F(\varphi)$. Q.E.D.

6. A variational problem with state constraints.

Let $f: R^n \times R^n \to R$ and $g: R^n \to R$ be given locally Lipschitz functions, and let two points x_0, x_1 of R^n be specified. Denoting the set of absolutely continuous functions $x: [0,1] \to R^n$ whose derivatives belong to $L^\infty([0,1], R^n)$ by A, we consider the following problem in the calculus of variations:

$$\text{minimize} \int_0^1 f(x(t), \dot{x}(t))dt \quad \text{over} \quad A\,,$$

subject to

$$x(0) = x_0, \quad x(1) = x_1 \tag{6.1}$$

and

$$g(z(t)) \leq 0, \quad t \in [0,1]\;. \tag{6.2}$$

As usual, we call weak local minimum any z in A which solves the above problem relative to the elements x of A whose derivative $\dot{x}$ lies in some uniform neighborhood of $\dot{z}$.

Theorem 4. Let z provide a weak local minimum for the problem described above, and suppose that whenever $g(z(t)) = 0$, we have $0 \notin \partial g(z(t))$. Then there exist an absolutely continuous function $p: [0,1] \to R^n$, a (nonnegative) Radon measure m supported on the set $\{t: g(z(t)) = 0\}$, and a measurable function $\gamma: [0,1] \to R^n$ satisfying $\gamma(t) \in \partial g(z(t))$ m-a.e., such that (we denote $\int_{(0,t)}$ by $\int_0^t$)

$$(6.3) \qquad (\dot{p}(t),\ p(t) + \int_0^t \gamma(s)\, m(ds)) \in \partial f(z(t), \dot{z}(t)) \quad \text{a.e.}$$

Remark. One can always reduce the case of multiple constraints $g^i(x(t)) \leq 0$ $(i = 1, 2, \ldots n)$ to the one treated above by defining $g(z) = \max_{1 \leq i \leq n} g^i(x)$. If each g^i is locally Lipschitz, then so is g. Furthermore, in the case when each g_i is C^1, the condition that 0 not belong to $\partial g(z(t))$ is equivalent (in light of Theorem 1) to the requirement that 0 not belong to the convex hull of the points $Dg^i(z(t))$ $(i = 1, 2, \ldots n)$. This is significantly weaker than the usual requirement that the vectors $Dg^i(z(t))$ be linearly independent.

Proof: Let $X = R^n \times R^n$, and define S to be the following closed subspace of $L^\infty([0,1], X)$:

$$S = \{(x, y) \in C([0,1], R^n) \times L^\infty([0,1], R^n):\ x(t) = \int_0^t y(s)ds \quad \text{a.e.}\}\ .$$

We define F and G, functions from S to R, via

$$F(x, y) = \int_0^1 f(z(t) + x(t),\ \dot{z}(t) + y(t))dt\ ,$$

$$G(x, y) = \max_{0 \leq t \leq 1} g(z(t) + x(t))\ ,$$

and we define $H_i: S \to R$ $(i = 1, 2, \ldots n)$ by

$$H_i(x, y) = x_i(1)$$

(the i^{th} coordinate of the vector $x(1)$).

It follows that the element $(0, 0)$ of S is a local minimum on S for $F(x, y)$ subject to $G(x, y) \leq 0$ and $H_i(x, y) = 0$ $(i = 1, 2, \ldots n)$. Since all these functionals are Lipschitz, we may apply [9, Theorem 1] to conclude that there exist numbers α, β, δ_i $(i = 1, 2, \ldots n)$ not all zero such that

$$\alpha \geq 0, \quad \beta \geq 0, \quad \beta G(0, 0) = 0 \quad ,$$

and such that

$$0 \in \alpha \partial F(0, 0) + \beta \partial G(0, 0) + \sum \delta_i \, \partial H_i(0, 0) \quad . \tag{6.4}$$

The generalized gradient ∂F is described by Theorem 3 (§5). In the case of G, note that

$$G(x, y) = \max_{0 \leq t \leq 1} \hat{g}(t, \pi(x, y)) \quad ,$$

where $\pi: S \to C([0, 1], R^n)$ is the "projection"

$$\pi(x, y) = x \quad ,$$

and where $\hat{g}: [0, 1] \times C([0, 1], R^n) \to R$ is the "evaluation"

$$\hat{g}(t, x) = g(x(t)) \quad .$$

We invoke §1 articles 13, 17 and §3 Theorem 2 to deduce that for any ζ in $\partial G(0, 0)$ there exists a probability Radon measure μ supported on the set

$$\{t: \; G(0, 0) = g(t, z(t))\}$$

and $\gamma(t)$ belonging to $\partial g(z(t))$ μ-a.e. such that, for all (x, y) in S,

$$\langle (x,y), \zeta \rangle = \int_{[0,1]} \langle x(t), \gamma(t) \rangle \, \mu(dt) \ .$$

After a simpler study of ∂H, we find from (6.4) that for certain integrable functions ζ_1, ζ_2 such that

$$(\zeta_1(t), \ \zeta_2(t)) \in \partial f(z(t), \ \dot{z}(t)) \quad \text{a.e.} \ ,$$

for certain μ and γ as described above, and for the vector $v = [\delta_1, \ldots, \delta_n]$ in R^n, we have

$$(6.5) \quad \alpha \int_0^1 \{\zeta_1(t) \cdot x(t) + \zeta_2(t) \cdot \dot{x}(t)\} dt + \beta \int_{[0,1]} \gamma(t) \cdot x(t) \, \mu(dt) + v \cdot x(1) = 0$$

whenever $x \colon [0,1] \to R^n$ is absolutely continuous and $x(0) = 0$.

A standard argument from the calculus of variations (see [16, p. 50]) obtains the following from (6.5):

$$(6.6) \qquad \alpha\zeta_2(t) = c + \alpha \int_0^t \zeta_1(s) ds + \beta \int_0^t \gamma(s) \, \mu(ds) \qquad \text{a.e.} \ ,$$

for some constant c in R^n.

Suppose that α is 0. It follows from (6.5) that β is strictly positive (for $\beta = 0$ would imply $v \neq 0$), and hence from (6.6) that $\int_0^t \gamma(s) \, \mu(ds)$ is constant. Also, since $\beta > 0$ we have $G(0,0) = 0$, which implies that on the support of μ, γ is never 0. This is a contradiction, so that we can assume $\alpha = 1$. If we set

$$m = \beta\mu, \quad p(t) = c + \int_0^t \zeta_1(s)ds \quad ,$$

the conclusion of the theorem is seen to follow (that the support of m lies in the stated set is immediate if $G(0, 0) = 0$; but if $G(0, 0) < 0$ then $\beta = 0$ and $m = 0$). Q. E. D.

REFERENCES

[1] J. P. Aubin, A. Cellina, J. Nohel, Monotone trajectories of multivalued dynamical systems, to appear.

[2] N. Bourbaki, Variétés Différentielles et Analytiques, Hermann, Paris.

[3] F. H. Clarke, Necessary conditions for nonsmooth problems in optimal control and the calculus of variations, thesis, University of Washington 1973 (Supervisor: R. T. Rockafellar).

[4] ___________, Generalized gradients and applications, Trans. Amer. Math. Soc., 205(1975), 247-262.

[5] ___________, The Euler-Lagrange differential inclusion, J. Differential Equations, 19(1975), 80-90.

[6] ___________, Necessary conditions for a general control problem, in Calculus of Variations and Control Theory, (edited by D. L. Russell), Mathematics Research Center (University of Wisconsin-Madison) Pub. No. 36, Academic Press, N.Y. (1976).

[7] ___________, The generalized problem of Bolza, SIAM J. Control Optimization, 14(1976), 682-699.

[8] ___________, On the inverse function theorem, Pacific J. Math., 62(1976).

[9] ___________, A new approach to Lagrange multipliers, Math. Operations Research, 1(1976), 165-174.

[10] F. H. Clarke, The maximum principle under minimal hypotheses, SIAM J. Control Optimization, 14(1976).

[11] ____________, Inequality constraints in the calculus of variations, to appear.

[12] N. Dunford, J. T. Schwartz, Linear Operators (Part I), Wiley Interscience, N.Y. (1957).

[13] A. Feuer, Minimizing well-behaved functions, in Proceedings of the Twelfth Annual Allerton Conference on Circuit and System Theory, (Illinois) (1974).

[14] A. A. Goldstein, Optimization of Lipschitz continuous functions, to appear.

[15] H. Halkin, Mathematical programming without differentiability, in Calculus of Variations and Control Theory, (Edited by D. L. Russell), Mathematics Research Center (University of Wisconsin-Madison) Pub. No. 36, Academic Press, N.Y. (1976).

[16] M. R. Hestenes, Calculus of Variations and Optimal Control Theory, Wiley, N.Y. (1966).

[17] E. Hewitt, K. Stromberg, Real and Abstract Analysis, Springer-Verlag, N.Y. (1965).

[18] J. B. Hiriart-Urruty, thesis, Université de Clermont (1976).

[19] A. D. Ioffe, V. L. Levin, Subdifferentials of convex functions, Trans. Moscow Math. Soc., 26(1972), 1-72 (English translation).

[20] G. Lebourg, Comptes Rendus de l'Académie des Sciences de Paris, November 10, 1975.

[21] __________, Generic smoothness properties of locally Lipschitzian real-valued functions defined on open subsets of infinite dimensional topological vector spaces, to appear.

[22] R. Mifflin, An algorithm for nonsmooth optimization, to appear.

[23] B. S. Morduhovic, to appear.

[24] B. H. Pourciau, Analysis and optimization of Lipschitz continuous mappings, J. Optimization Theory Appl., to appear.

[25] B. N. Pshenichnyi, Necessary Conditions for an Extremum, Marcel Dekker, N.Y. (1971).

[26] R. T. Rockafellar, Conjugate Duality and Optimization, SIAM Publications, Philadelphia (1974).

[27] T. H. Sweetser, A minimal set-valued strong derivative for vector valued Lipschitz functions, to appear.

[28] L. Thibault, Quelques propriétés des sous-differentiels de fonctions réelles localement lipschitziennes définies sur un espace de Banach séparable, Comptes Rendus Acad. Sci. Paris, 282(1976), 507-510.

[29] J. Warga, Derivate containers, inverse functions and controllability, in Calculus of Variations and Control Theory, (Edited by D. L. Russell) Mathematics Research Center (University of Wisconsin-Madison) Pub. No. 36, Academic Press, N.Y. (1976).

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
#1687

2. GOVT ACCESSION NO.

3. ~~RECIPIENT'S CATALOG NUMBER~~
9 Technical

4. TITLE (and Subtitle)
6 GENERALIZED GRADIENTS OF LIPSCHITZ FUNCTIONALS,

5. TYPE OF REPORT & PERIOD COVERED
Summary Report, ~~no specific reporting period~~

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
10 Frank H. Clarke

8. CONTRACT OR GRANT NUMBER(s)
15 DAAG29-75-C-0024

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Mathematics Research Center, University of Wisconsin
610 Walnut Street
Madison, Wisconsin 53706

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, North Carolina 27709

12. ~~REPORT DATE~~
11 October 1976

13. ~~NUMBER OF PAGES~~
30

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
12 34p.

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public ~~release; distribution unlimited.~~
14 MRC-TSR-1687

17. DISTRIBUTION STATEMENT ~~(of the abstract entered in Block 20, if different from Report)~~

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Locally Lipschitz functions, subdifferential, generalized derivatives, generalized gradients.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of this article is threefold: (i) to present in a unified fashion the theory of generalized gradients, whose elements are at present scattered in various sources; (ii) to give an account of the ways in which the theory has been applied; (iii) to prove new results concerning generalized gradients of summation functionals, pointwise maxima, and integral functionals on subspaces of L^∞. L(INFINITY)

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 221200 1B